quality of working life

EVALUATION AND MEASUREMENT

Quality of Working Life

Evaluation and Measurement

D.A. Ondrack
M.G. Evans
Faculty of Management Studies
University of Toronto

Labour Canada Travail Canada

Published by Authority of the Honourable Gerald A. Regan,
Minister of Labour, Government of Canada.

The opinions expressed in this booklet do not necessarily
reflect the views or policies of Labour Canada.

January 1981

Également disponible en français sous le titre

Évaluation et mesure de la qualité de la vie au travail

Available from:

Quality of Working Life
Employment Relations and
Conditions of Work
Labour Canada
Ottawa, Ontario
K1A 0J2
(819) 997-2861

© Minister of Supply and Services Canada 1981

Cat. No. L82-38/1981E

ISBN 0-662-11431-0

THE AUTHORS

Daniel Ondrack is Associate Professor of Organizational Behavior at the Faculty of Management Studies, University of Toronto and is also Professor of Personnel Administration with the Centre for Industrial Relations, University of Toronto. He is currently engaged in a joint research project with Professor Martin Evans on a series of case studies of Quality of Working Life Programs in the petrochemical industry.

From 1975-77 Professor Ondrack was a Visiting Professor of Organizational Behavior at the European Institute for Advanced Studies in Management. There he studied European experiments in Quality of Working Life and has produced a book on the subject which will appear shortly. He is a member of the Board of Directors of the Personnel Association of Ontario. He received his Ph.D. from the University of Michigan and his M.B.A. from the University of Toronto.

Martin G. Evans is Professor of Organizational Behavior at the Faculty of Management Studies, University of Toronto. His research interests focuses upon motivation in organizations, performance and job satisfaction. He has been involved in a number of studies on the evaluation of organization development programs, flexible working hours, Blake grid and supervisory training programs. He has published extensively in academic journals and is a member of the editorial board of the Academy of Management Journal. He received his M.Sc. from Manchester University and his Ph.D. from Yale University.

Acknowledgments

We should like to thank Carl Johnston, Air Canada; R.M. McLeod, Labour Canada; Jack Dougall, Dalhousie University; Jak Jabes, University of Ottawa; Harvey F. Kolodny, University of Toronto; Pierre-B. Lesage, Université de Montréal; N.D. Porter, Union of National Defence Employees; Denis Roy, Université de Montréal, Dieter Schachhuber, Public Service Alliance of Canada; and Terrence White, University of Alberta, for their thoughtful reading of the earlier drafts of this booklet and for their useful comments and suggestions. Needless to say, we were not able to incorporate all of the suggestions, so that responsibility for the final version rests with the authors.

D.A. Ondrack
M.G. Evans

Toronto, June 1980

PREFACE

Judging from the requests we have received, this third in our series of booklets on the subject "Quality of Working Life: Measurement and Evaluation" should find a receptive audience. We trust that our readers will find the present booklet as helpful as these issued earlier in the series "QWL: The Idea and the Application" and "QWL: Job Redesign and Sociotechnical Systems".

Many of us would agree on the importance of taking stock of where we are going, assessing what has been accomplished, learning from our experience and adjusting our future plans accordingly. This evaluation process is particularly relevant, I believe, for those of us involved in Quality of Working Life projects within organizations whether as managers, union officers, employees or practitioners. The present booklet has been prepared with this need in mind.

It is our hope that this booklet will help our readers acquire a better understanding of the evaluation process within the QWL context and provide useful suggestions that will assist organizations in the development and application of appropriate measurement and evaluation techniques suited to their needs. The effective evaluation of projects, it seems to me, is highly relevant both to their continuation and their more extensive diffusion within organizations.

Our series of booklets is designed to assist those involved with the practical application of QWL principles in Canada. This often involves radical adjustments in our thinking and working habits. It involves experimentation with new concepts and new ways of working together in a cooperative relationship. The experimental nature of projects suggests the need for careful and systematic measurement and evaluation and for changes in project design and administration in the light of experience, changes which may be necessary if project objectives are to be realized.

T.M. Eberlee,
Deputy Minister, Labour Canada

Contents

		Page
I.	Changing Organizations	1
II.	Why Evaluate?	3
III.	The Change Process: Start-up and Change-over Situations	6
IV.	Testing for the Implementation of a Program: The Manipulation Check (Formative Evaluation)	16
V.	Summative Evaluation and Analytical Techniques	17
VI.	Action Research and Formative Evaluation - A Feasible Approach	37
VII.	Conclusion	41

List of Tables and Figures

Table 1 - Evaluation of a Change-over Situation 8

Table 2 - Variables Affecting Success in the Last Two Phases of a QWL Program 23

Figure I - Stages of Start-up Situations 11

Figure II - Stages of Start-up and Types of Evaluation 12

Figures III-VIII Examples of Changes in productivity over time

Figure III 25

Figure IV 26

Figure V 26

Figure VI 31

Figure VII 33

Figure VIII 34

I. Changing Organizations

All programs of organizational change, have a common underlying process which can be divided into five stages: initiation; set-up; work situation analysis; design and implementation; evaluation of results. The major concern of this work is with this final step -- evaluation. However, activities essential to the evaluation stage have to be performed at each of the earlier stages of a program. The discussion of the implementation and evaluation stages that follow is geared for a Quality of Working Life Program which assumes management/union/employee collaboration in planning and implementation. The Labour Canada booklet defines Quality of Working Life as involving the redesign of jobs and their group and organizational contexts such that some balance between technical and social efficiency is achieved through a process that involves the participation of employees, managers, and (in some cases) unions in the redesign process.[1]

Stage 1 - Initiation

At this stage: (a) the need for change (or at least the need to explore whether or not change would be useful) is recognized, and (b) an attempt is made to ensure the commitment and participation of the interested parties (union, workers, management) to the change process.

Stage 2 - Set-up

At this stage of the project, an agreement is worked out which specifies: (a) what is to be done; (b) any constraints within which the program has to operate (such as particular clauses in the existing union contract, "trigger" levels of productivity or turnover at which a program might be terminated); and (c) the development of a joint steering committee designed to oversee the general progress of the change. In addition, the units of the organization that are to take part in the change program are specified. Finally, decisions are made as to which people are going to be involved in the Project Team which does problem diagnosis and works on the implementation of change.

Stage 3 - Work/Situation Analysis

Each work situation analysis is unique because each firm or indeed a department in a firm faces unique problems. However, the diagnostic process in a variety of change programs is usually similar. Diagnosis normally begins by identifying existing problems and by describing current work flow and the content of jobs performed by employees and their supervisors. A diagnosis also analyses whether or not the tasks are appropriately organized, co-ordinated and rewarded.

---

[1] C.F. Johnston, M. Alexander and J. Robin, Quality of Working Life: The Idea and Its Application, Ottawa, Labour Canada, 1978.

The people involved in the diagnosis may require additional training and access to outside consultants. For evaluation purposes, this diagnosis stage may provide initial data about the state of the organization prior to the change.

Stage 4 - Design and Implementation

This stage may well be the longest in time and involves a sequence of linked substages. The Project Team develops a program for change (redesigned jobs, a new payment system etc.) and begins to implement it. This may involve retraining of personnel and the addition of new information systems, as well as changing the kinds of jobs that people do or the relationships between employees and managers. At the beginning of this stage it is essential to set a series of goals to be achieved over the course of the change. Without these, any change project is likely to lose momentum as the pressures of everyday work reassert themselves. The timely completion of each goal provides useful data for the evaluative process.

Stage 5 - Evaluation and Diffusion

In this stage, the Steering Committee reviews the work of the project team, establishes whether or not change has occurred, and establishes whether or not the desired outcome has been achieved. With the benefits of appropriate data, the Steering Committee establishes: (a) whether or not the changes induced were responsible for bringing about the desired outcome; (b) the conditions that were conducive to establishing these results; and (c) decides whether other units in the organization will be sites for an extension of the change program. This is not to say that the program is to be adopted unmodified in the diffusion stage. As the Labour Canada QWL Booklet suggests:

> It should not be assumed that a second team will simply adopt the solutions of its predecessor. A new, almost equally complex analytical and creative process will be undertaken, and although at the end of the development period there may be substantial agreement on solutions to major problems, each project will put its own stamp on its work situation.

The cycle of set-up, diagnosis, implementation, and evaluation is to be repeated for each program, however, the diffusion should benefit from the learning generated by the previous programs.

While people in the project team have responsibility for evaluation, this does not necessarily mean that they have to carry out the evaluation. Some would argue that ideally, people responsible for evaluation should be different from those responsible for implementation but this may not always be practical or feasible except in very

large programs. An example of this can be found at Alcan[2] where one person from the Employee Relations Department acted as the technical consultant, the process consultant, and the evaluator.

The essential point is that concern for evaluation should occur almost simultaneously with concern for solving a problem or implementing a program. Persons involved in implementing a program should think of evaluation in terms of having to report results to a sceptical client. This does not mean that evaluation should only be concerned with positive results. Much can be learned about programs and implementation techniques from negative results. However, if the sceptical client is going to be convinced to allocate more resources to a program or authorize further programs, one has to make a convincing case. He will rarely be won over by good intentions or informal impressions. Consultants, workers, or managers all need the results of planned, systematic evaluation in order to profit from experience, solve problems, or obtain resources for other programs.

II. Why Evaluate?

More QWL programs are implemented than are implemented and evaluated. Such programs are often begun because someone thinks it is a good idea. When they are completed (or even worse when they peter out), people in the organization simply move on to other activities. At best, a highly informal evaluation might be done in the form of subjective impressions of how the program went, how much people enjoyed the program, and whether they seemed to "get anything" out of it. More often, the basic concern is to complete the program and to get started on something else. As a result, we often do not have a good idea about which programs are effective or under what conditions a particular program is effective. The state of affairs described above can be easily understood if evaluation is too difficult or not felt to be worthwhile. Thus, it is necessary to set out the case for evaluation before proceeding to discuss approaches to performing evaluation.

Basically there are three reasons for doing evaluations:

1. The Investment Rationale - A QWL program could be justified either as the "right" thing to do or as a means to more effective organization performance. If it is the "right" thing to do, then beliefs and values are the main rationale for a particular program. Although the reason for a QWL program is as intangible as someone's beliefs and values, an investment must still be made to achieve these goals. Time and money must be spent to do the necessary planning and training, to take people off the job, to make the change-over from one

---

[2]See pp. 14-15. The quality of worklife experiences of the Aluminum Company of Canada at Arvida, Quebec will be referred to extensively in this booklet to illustrate stages and techniques of QWL evaluation.

work system to another. Thus even in cases where someone insists, "we have to work in this way because it's the right way to work", it is still desirable to do an evaluation to see what it actually costs to do things "the right way."

However, the more common and compelling argument for implementing a QWL program is the belief that it will result in more effective organization performance. In this case, the investment in training, etc., should be equated with the payoff, i.e., comparing the results before and after the program. If after investigation, the difference in results does not compensate for the program cost, then one has convincing evidence for changing or dropping it. On the other hand, if it turns out that the results consistently exceed the costs, then one can conclude that the programs should be continued or even expanded.

2. The Implementation Effectiveness Rationale - Implementing a QWL program requires the use of several Organization Design (OD)[3] techniques. The techniques can range from a large-scale, organization development effort, to some rather basic training. While highly experienced professionals may have accumulated an array of effective implementation methods over time, they, like most of us, are concerned with learning how to do things better. We all need some way to profit from our experience so that we can become more effective over time.

Planned, systematic evaluation can aid this learning process because data can be obtained on results of programs and related techniques. It is also possible to experiment with different programs or techniques where one can count on getting valid feedback on the results. Without evaluation, both organization and those implementing the program stand to lose valuable opportunities to learn from experience and to increase effectiveness.

3. The Legislation Rationale - A third reason for evaluation which has recently gained prominence, is the need to satisfy regulatory requirements. In Ontario for example, recent legislation on occupational health and safety requires that safety co-ordinators monitor organizational safety programs. This is a case of the evaluator role being legislated into existence.

There are thus compelling reasons to evaluate, to plan in advance for evaluation and to include evaluation procedures as an integral part of QWL program implementations. Planning in advance

---

[3]OD, Organizational Design or Development is the name given to the application of behaviour science techniques to the improvement of the total organization, through planned, systematic, long-range efforts and focussing on human interactions, usually assisted by a change agent or catalyst. For further reading on OD, see the references to French and Bell 1978, Beckhard and Harris 1977, and Nadler 1977.

for evaluation requires thinking about the objectives of a program and the kind of data that should be obtained in order to determine if objectives have been achieved. An example of the objectives in one case was:

(a) one objective of this program is to reduce time-loss accidents among production staff,

(b) successful completion of this objective will be recognized when accident rates fall to and remain at X per cent.

The effort of going through such an exercise before a program is implemented will enforce more deliberate, more careful, and more systematic thinking about programs. Of course, in the initial stages there may be a lot of ambiguity and uncertainty about the problems facing the organization and the solutions available. We are suggesting that once an organizational diagnosis has been completed and the decision made to implement changes, then the milestones for evaluating the success of the stages should be included as part of the implementation plan.

Approaches to Evaluation

Another reason for the failure to evaluate may simply be intimidation at the rigours of the approach used in most books and articles on program evaluation, namely, the scientific method. This method is designed to add to the body of knowledge and, generally, the only acceptable additions derive from a formal experiment. An experiment is designed to change one or a few factors in a situation while all other factors are held constant for a sufficient length of time to observe the eventual impact of change. In even a simple experimental design, a test group of workers is distinguished from a control group; the latter will not be subject to the experimental change. No changes in the work situation of either group may occur for the duration of the experiment. Both groups are pretested, and if the results of the test at the end of the experiment show that a change did occur in the test group and not in the control group, then one can have reasonable confidence that the experimental change had something to do with the behaviour change observed in the test group.[4]

For example, if one wanted to test the effect of room temperature on worker behaviour, one could select two rooms with largely similar workers, raise the temperature in one room and leave the thermostat unchanged in the other. Based on some prior experience with the relationship between hot temperatures and behaviour, one might even make a prediction of what sort of effects one would expect to see from the experimental change. This is the same as forming hypotheses about the experiment, i.e., "the workers in the hot room

[4]Agnew and Pyke, The Science Game, 1969, For an easy-to-read elaboration on scientific research.

will show signs of discomfort, become drowsy, etc. The workers in the normal room will show no change." Then after a period of time when the test room is observed to be full of uncomfortable, drowsy workers and the other room has no change, one has confirmed the hypothesis and thereby demonstrated a relationship between room temperature and worker behaviour.

This is a very simplistic description of the scientific method. For most cases in an organizational setting it is unfortunately not possible to achieve the conditions to conduct a true experiment. For example, if the workers in the room saw you fiddling with the thermostat, they might have responded to what they saw just as much as the actual temperature. This is a problem of people responding to the fact that an experiment is being conducted (Hawthorne effect)[5], as much as to the changes introduced by the experiment. Secondly, all other factors in an experiment are to remain unchanged during the course of the experiment. What if someone opens a window; walks up to the thermostat and changes it; walks into the other room and discovers a temperature difference, etc.? If you tried to prevent any of these actions in order to hold all other factors constant, you might be behaving so oddly that your strange behaviour (i.e., "Dont't touch that thermostat") contaminates the experiment. Finally, there is the ethical question of unilaterally and secretly imposing changes in working conditions without explanation or consultation beforehand with the workers to be affected by the change.

Our approach to evaluation therefore acknowledges that the requirements of strict scientific procedure is likely to be frustrating in a real-life organizational setting. Indeed, the object of those who initiate QWL programs is rarely to add to the body of scientific knowledge and theory and it is probably not appropriate to aspire to scientific rigour when trying to evaluate such programs. Instead a compromise approach is required, one that gives reasonably valid feedback about the results of programs and techniques and yet is operationally feasible.

III. The Change Process: Start-up and Change-over Situations

Behaviour of people in an organization may fluctuate from day-to-day, but largely falls into general patterns as long as no major disruptions occur in the organization or its environment. Given a certain amount of stability in an organization, then behaviour will tend to settle into routines. Thus, a QWL program initiated in an existing situation (change-over) will present problems different from those in a new organization (start-up situation).

---

[5] A term derived from research conducted at the Hawthorne works of the Western Electric Company in the 1930s. In the Hawthorne effect worker behaviour changes simply because unusual attention is paid to their behaviour during an experimental program and the behaviour itself has little correspondence to the nature of the experiment.

In change-over situations a conscious attempt is made to change the routines of behaviour in the organization to give up the old ways and adopt new ways. The techniques for implementing such changes can range from coercion/power strategies to sophisticated organization development (OD) strategies. A common phenomenon in change-over situations is that behaviour will likely slide back to the old routines soon after the monitoring of the change has ceased. The reason for this regression is that the forces in the social environment surrounding the work behaviour were more strongly in favour of the old behaviour than the new. Without a program of complementary changes in auxiliary support systems in the organization (i.e., compensation policies, performance standards, collective agreement), the change-over behaviour is very likely to revert back to the old ways.

What this means for evaluation, is that the project team must look at the "before" behaviour of a work system in question, plus the "before" social environment surrounding the work system. Thus, a multi-dimension approach is required in order to make realistic and practical evaluations of change-over situations. Table 1 below illustrates how this might be done.

From this figure we can see that two parallel evaluation processes are being used. On the one hand (B), an evaluation is being done of the work site where it is hoped that a QWL program will effect changes. On the other hand (A), evaluation is being done of at least some of the principal factors in the work site environment. Examination of work site changes might reveal that some positive changes had occurred and from that data we might conclude that a successful change-over had occurred. However, if we find that no parallel changes had occured in the work environment, we can probably be justified in concluding that the change-over will be temporary. Some theoreticians of evaluation have even suggested that a work site change-over without corresponding environmental change is evidence only of a positive response to being studied, the classic Hawthorne effect and not even a temporary change-over, i.e., the behaviour will regress soon after monitoring ceases.

Table 1: Evaluation of a Change-over Situation

A - Evaluation of Work Environment Changes

Before Implementation of Program	During Implementation	After Implementation
1. Foremen's Attitudes and Supervisory Practices	1. Monitor Foremen Training Program	1. Foremen's Attitudes and Supervisory Practices
2. Incentive System	2. Monitor Redesign of Incentive System	2. Incentive System
3. Performance Standards	3. Monitor Redesign of Performance Standards	3. Performance Standards

B - Evaluation of Work Site Changes

Work Site Data Before Implementation	Work Site Data During Implementation	Work Site Data After Implementation
1. Quality Record	1. Monitor Quality Record	1. Quality Record
2. Production Volume	2. Monitor Production Volume	2. Production Volume
3. Waste Levels	3. Monitor Waste Levels	3. Waste Levels

A key requirement of the parallel or multi-dimensional evaluation method for change-over programs is that one should attempt a parallel change strategy. The changes in foremen attitudes and supervisory practices, changes in incentive schemes, etc., are not likely to take place unless a program exists to change these factors as well as behaviour at the work site. Thus, implementation of a change-over program can be greatly enhanced if a multi-dimensional change strategy is used. In many ways, a change-over QWL program is the most difficult type to achieve due to the considerable behavioural inertia found in many organizations. Planning for a comprehensive or multi-dimensional evaluation strategy may have a significant side-effect of aiding the planning for a more sophisticated implementation strategy for a QWL program.

An example of a change-over program is the Aluminum Company of Canada (Alcan's) QWL experiment in part of its smelter in Arvida, Quebec from 1968-1971. While the Arvida site employed 6000 people at that time (4500 hourly), the experiment involved only one department of an aluminum casting plant which had 47 hourly paid workers. There were an average of 16 workers per foreman; the foremen were usually very rushed; worked by giving orders and had little time for discussing problems with their workers. The main impetus for the change program came from a superintendent who wanted to improve performance of the department and provide a means for the workers to have more satisfying work.

Over the years, management had introduced a series of technological changes which raised production and demanded higher levels of skills from the workers. At the same time, the livelihood of the workers was threatened because the technological changes reduced demands for manpower. Promotion was determined by seniority alone. The effect was that promotion was a payoff for having endured a sequence of poor jobs rather than being an incentive to learn new skills or to accept responsibility.

To redress some of the problems of promotion and salary, as well as to continue to achieve gains in productivity, a "sheltered experiment" was proposed where a new type of organization could be tested without any penalty to the employees, and without involving the whole plant in a possibly disruptive venture. A full description of the objectives of the experiment and the means and stages of implementation is available in Jones (1975) description of the case. Further references to aspects of the case will be used to illustrate various points in the balance of this handbook.

The formation of work teams was the main change implemented in the Alcan experiment. Each team member learned and rotated between the various jobs assigned to the team. Training was done during regular working hours and the workers were paid an increment when they passed a test of knowledge and skills for each new occupation. The foreman's role was also modified to focus on planning and co-ordinating. Extensive pre-change training was undertaken for both foremen and workers. These changes in working arrangements were expected to

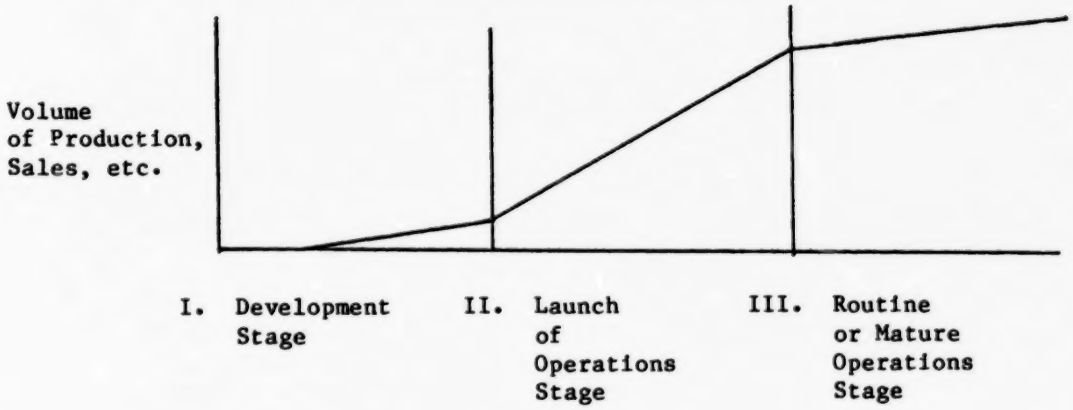
result in changes in employee attitudes and increased job satisfaction. An interesting aspect was agreement by the parties to try certain work arrangements without knowing all the details prior to the change, i.e., the specifics would be known as they developed through the life of the experiment. The evaluation that took place at the end of the experiment was done jointly by the union and the company. Agreement by each party was required before the changes could become permanent. Based on the evaluation results, the parties agreed to implement a modified but basically similar system for all casting department employees of the plant (over 400 employees).

Start-up programs in contrast to change-over situations, typically involve a new facility, e.g., the construction of a new plant, the creation of a new department or the starting of a new business. Of course, there is no such thing as a pure start-up situation as every person brought in to work in the new situation carries with them work habits and behaviour patterns acquired through prior work experience. Few, however, will have had experience in working together. In addition, they will not necessarily be familiar with the technology of the new operation. Thus, a start-up situation presents great opportunities for creating a new social and technical work environment that will encourage the development of desirable behaviour patterns. An added advantage is that there is much less behavioural inertia to overcome than that inherent in the change-over situation.

However, a start-up situation poses at least two special problems for evaluation. The most immediate problem is the lack of an ongoing behaviour pattern from which to derive pre-program measures of work behaviour, attitudes, performance levels and so on. Similarly, there are not likely to be any opportunities to compare program results of a test group with control groups. (It is not likely in a start-up situation that an organization would implement a QWL program with one group of employees and not the other.) For these reasons, evaluation in a start-up situation cannot make use of the classic "before and after" comparisons associated with a change-over situation.

The second special problem for evaluation of a start-up situation is a problem similar in some ways to the Hawthorne effect. People's behaviour can be affected as much by the fact that a start-up is occurring as they are affected by a QWL program occurring simultaneously with the start-up. The reason this can happen is that a start-up can go through three stages of development, each stage presenting a different environmental situation for the employees involved. The stages are shown in Figure I below.

Figure I: Stage of Start-up Situations



In Stage I, the type of work in which people are engaged is mostly training and preparation. They have little or no opportunity to work together in the actual operation of the new system. Some will be waiting to see how the system really operates before they make a judgment on whether they like working there or not. Some will do the opposite and assume that the organization climate during the development stage will be the same during the later stages. For these reasons, any evaluations conducted during the development stage can only be tentative. The situation is artificial and the responses must be regarded as artificial as well.

In Stage II, there is often a highly charged atmosphere with considerable excitement and the dual pressures of achieving productive operations while maintaining the integrity of the QWL program. The situation can be highly satisfying to those who are motivated by challenge and problem-solving but threatening to those with low tolerance for ambiguity and uncertainty. There can be great pressure to get the new facility running fast and to take corrective action quickly. Supervisors and managers may tend to call the shots themselves without "wasting time" on participative discussions with subordinates. If this is not checked by the monitoring system, subordinates may feel deceived and disappointed by the failure to meet the expectations created in the training programs.

Similarly, some people are especially motivated by the dynamic conditions during a launch stage and will respond positively to all evaluations conducted during this phase. Later when things slow down and operations become more routine, they may actually leave the organization to seek launch stage work elsewhere. It is well known in the construction industry that some persons are "start-up gypsies", travelling from one new site to another in search of the

adventure of the launch stage. This behaviour can show up in a sudden rise in turnover about the time that a start-up moves from Stage II to Stage III. An evaluation conducted without awareness of this phenomenon could be misleading.

In Stage III most of the settling-down and shake-outs have occurred. The social environment of the organization is becoming established and work behaviour is settling into routines. It is only during this third stage that we can begin to speak of "normal" behaviour in the organization because the start-up itself has ceased to be a major factor in influencing behaviour. Figure II below illustrates the relationship between stages of start-up and types of evaluation.

Figure II: Stages of Start-up and Types of Evaluation

I. Development Stage	II. Launch Stage	III. Routine Stage
Situation →		
Training Activity, Expectations Created, Hypothetical Work and Artificial Behaviour	Uncertainty, Problem-solving Challenge, Learning, Excitement, QWL Programs are Latent	Establishing Routines and Procedures, Typical Behaviour Patterns Emerge as QWL Programs Become Reality.
Evaluation Looks at: →		
Training Content, Logical Consistency of Programs and Support Systems	Monitoring for Deviations from Plans; Achievement of Interim Objectives	Target Achievements in Climate Attitudes, Behaviours, Production Criteria

As the start-up model implies, the project team must have considerable patience during the evaluation process. Several interim evaluations can be conducted during the whole start-up process but these can only deliver tentative results. For a more conclusive evaluation of a start-up program, one might have to be prepared to wait two or three years.

The QWL project at Petrosar is an example of a start-up situation which attempted to create a distinctive plant culture and set of working arrangements where none had existed before. Rather than starting with one experimental group as had been done at Alcan, and later expanding to the rest of the department, the Petrosar project team started with the first people hired. All new personnel

had to be indoctrinated and trained in the system as they came on board. The absence of past practices gave Petrosar freedom of design for the new plant but also made careful planning for hiring, indoctrination and training of vital importance.

The Petrosar plant uses crude oil to produce ethylene which in turn is used by other chemical plants to produce plastic products. Started in November 1977 with a work force of 500, it now has approximately 700 employees. The basic objective of the QWL program was to involve workers to a large degree in the decision-making actions and responsibilities that would normally be done by supervisors. Workers are trained in several different job skills and wage progression is based on number of skills learned. All are organized into largely autonomous shift teams and they must decide for themselves who is to do what, when, and for how long. Workers are also involved in several other types of decisions and problems through their participation in standing committees and project teams. A joint standing committee periodically collects data on workers attitudes and job concerns. If problems are perceived, it is the responsibility of the committees to take action. In general, the results of the program and the production levels achieved so far have been satisfactory.

It was earlier stated that evaluations of start-ups were more difficult because data for a before-after comparison of program results are not available. Yet the situation is not completely hopeless in that it may be possible to synthesize a before-after study or at least attempt some form of comparative evaluation. Some approaches are listed below.

1. Progress Toward Goals - If the planners of the start-up have done their homework and have identified program goals and assessment criteria, then evaluation can be done simply by measuring progress toward the goals. For example, the goal may be to create working conditions so as to reduce absenteeism, turnover and employee lateness. Where such criteria can be specified in advance, it is a relatively simple matter to do comparisons and monitoring in terms of progress toward the desired goal.

Similarly, more general objectives may be stated such as, "an objective of this program will be to establish a work climate where workers will feel a strong sense of responsibility, participation, and satisfaction with their work." To evaluate achievement of this objective, one might have to develop a questionnaire or interview schedule and try to obtain data on how workers actually feel about their sense of responsibility, participation, etc. Such a questionnaire might have an item like this:

No. 3 "I have a strong sense of meaningful participation in decisions affecting my work."

1	2	3	4	5
Disagree	Somewhat Disagree	Uncertain	Somewhat Agree	Agree

A series of replies toward the 4 or 5 end of the scale will indicate that workers are responding in a manner consistent with the goals of the program. Of course, the first time such a questionnaire is used, one doesn't know how workers felt prior to this situation but at least one could verify that current responses are consistent with program goals. Over time a series of such questionnaires can monitor progress toward the goals. This "time-series" approach to evaluation, where the early questionnaire results provide the "before" data provides a basis for comparative studies.

In the Petrosar case, surveys of organization climate, employees' attitudes, progress in training and comprehension of company policies have been used regularly for over three years. All data is summarized and fed back to workers and supervisors for discussion at team meetings. The data is compared to target objectives and trends over time are examined.

2. Reference Group Comparisons - A large number of personnel questionnaires developed by researchers and by commercial distributors provide reference group information along with the questionnaire. For example, a job satisfaction questionnaire might provide representative group scores on the questionnaire for different types of occupational groups, and some of these may be comparable to the occupational groups in the particular organization where one is attempting to evaluate a program. There is always a danger that the reference groups are not comparable, but even here one has at least some idea of how others have responded to the same questionnaire. Needless to say, this type of comparison can only be done using a standard, unmodified version of someone else's questionnaire, whether obtained from some commercial source or from some published research study. A commonly used procedure is to administer the Likert Organizational Climate questionnaires to employees to see how their perceptions of leadership, etc., compare to the standards in the Likert questionnaire.[6]

Reference group comparisons can also be done using statistics on employee behaviour, production records, etc. from the relevant industry without having to use a questionnaire. Certain turnover levels may be typical for a local labour market or industry. Average plant start-up times, or rates of capacity utilization may be known. These can all provide a basis for comparing progress. Depending upon competitiveness and secrecy in a given industry, some of these data may be difficult to obtain. At the very least in a start-up situation one can always compare operations in the new organization with operations in the older parts of the parent organization. Even if these latter operations are not strictly comparable to the new operation, it is still valid to compare indices of effectiveness in the older operations. For example, if production units per employee in an old plant are 100 units/day per worker and 200 units/day per worker in the new plant, this is going to tell you a lot about how well the new plant is doing, (especially if a goal for the new plant was 150 units/day per worker).

---

[6]See Likert, R., The Human Organization, 1967.

3. Personal Assessments - Is there a role for personal assessments in evaluating start-ups? Obviously this can be the least reliable and least valid method, subject to all sorts of bias and error. Yet, a qualified observer, one who knows a situation well and who can make intelligent judgments on program results, can still make meaningful assessments for evaluation purposes. Personnel observations may often be required in start-up situations where no comparative norms or data from other situations are available to provide a basis for comparison. For example, the results of a brand new operation with no previous corporate history cannot be compared with the results achieved in older parts of the parent organization. If no well-established questionnaires or attitude surveys, relevant to the goals of the new operation are available, there will be no opportunity to compare present survey results with some previous survey results. In this situation, comparisons might be made with survey responses from another industry or society. For example, a new petro-chemical plant in a remote northern location probably would not have a similar plant in a similar location with which to compare its turnover and absenteeism figures. Instead they must use reference data either from other northern industry (mining, logging) or from other petro-chemical plants (mostly located in southern, industrialized areas).

Personal assessments can be greatly enhanced if pre-planning for the QWL program has established goals for the program, or if the persons involved can compare present experience with former experience. The difference between personal assessments and other evaluations is that results are not measured. Instead, the observer senses or gets a feel for the results, either by talking to participants, observing their behaviour, or looking for indirect indicators. Data may be gathered from informal conversations with samples of workers (or their supervisors), observing social interactions on or off the job, reading graffiti on walls, types of grievances, comments in suggestion boxes and so on. Based on such personal assessment company officials at Petrosar, for example report an almost complete absence of petty politics, gamesmanship and conflict between employees and supervisors. The data gathered in this illustration are not systematic, probably not random, don't identify or define variables, and are not measurable in any quantifiable sense. An overall impression is formed by the accumulation of bits of data into a mosaic or overall impression of results or progress. Personal observations for some observers can probably be almost as accurate and valid as careful, quantifiable measurement processes, but for others the process can be hopelessly invalid or inaccurate. Thus, we strongly recommend that all personal observations be supplemented by measurement-derived data whenever possible, or even better, that measurement data be the foundation for evaluations and that personal observations be used only to supplement the primary evaluations.

There are two main categories of evaluation - Summative and Formative. Summative evaluation is designed to answer the questions:

> What was the overall result of the QWL program? Did it have the desired effect? Implied here is the additional question of whether or not the program got implemented as designed.
>
> Is it worth implementing the program in other segments of the organization or in other organizations?

Formative evaluation is designed to answer a set of rather different questions:

> Did the planned change actually get implemented as planned?
> What actually happened during the change?
> Under what conditions did the change occur?
> What can be done to modify the implementation of the change to ensure that the desired outcomes do in fact occur?

IV. Testing for the Implementation of a Program: The Manipulation Check (Formative Evaluation)

The first stage in carrying out any kind of evaluation is to ascertain whether or not the proposed changes have in fact been implemented. A number of change attempts, i.e., job design, the creation of autonomous groups, even flexible hours, fail to meet this criterion. The new system developed by the project team often never really does get adopted in practice. If this occurs, one may draw a false inference that the change has failed to have its expected impact. In this case, the failure is due to the lack of implementation. A good example of such a failure is recounted by two Yale University action researchers in their attempt to introduce job enrichment (increased skill variety, task identity, feedback and autonomy) in the stock transfer department of a bank. There were a number of implementation problems including the failure of a new computerized information processing system and the lack of employee training for the new tasks which the employees were to perform. As a result, jobs did not change significantly and scores on a diagnostic questionnaire showed no before and after differences.

How can this check on implementation be made?

1. A clear specification should be made of the new behaviour, job design, working relationships and organizational structures that will constitute the change. Normally, this specification will occur after the organization has undertaken its own diagnosis and will culminate in the fourth stage of a QWL program: that of job/work design.

2. Appropriate indicators of these variables should be developed. If possible, objective events or behaviour should be monitored and recorded: Are meetings taking place? Are people coming to work at different times? Are some people varying their arrival and departure times? Is the task increased in scope and significance? Observation and analysis of routine organizational records can be used to determine if such changes have taken place. In other cases, reports by the people involved may be the only data available. In our Alcan example, the researchers report data based upon ordinary organization records about whether or not the program "took" or not. Prior to the introduction of the experiment 91 per cent of the employees knew one or two jobs, no one knew more than three; after the experiment was over only 8 per cent knew one or two jobs and 75 per cent knew five or six jobs. This gives a clear indication that one of the major purposes of the program - the development of a multi-skilled work force - was being achieved.

In addition, it may be useful to know the kind of factors that influence whether implementation has occurred at the first stage of a change program.

1. The relationship between management and workers: whether this was characterized by trust or mistrust; whether each party was united in its goals or whether the various management groups had different expectations for the projects and whether divisions occurred among workers, the local union and the union headquarters.

2. The role played by the project team: was it dominated by the union, management, or the change agent; was the team seen to be operating effectively and did employees and managers feel that they could influence team decisions; what was the degree of participation in the workings and decisions of the project team.

3. The behaviour of the group implementing the change: whether they were receptive to ideas, helpful or pushy.

4. The clarity of the change program and the extent to which managers and employees understood and accepted the program. More specifically: whether the program was fair, whether things were moving too quickly or too slowly, and whether employees evaluated the overall impact of the program favorably.

This set of measures enables us to answer why there has been resistance to a program and perhaps even failure to adopt the program.

V. Summative Evaluation and Analytical Techniques

Summative evaluation is designed to answer the question "What has been the overall effectiveness of an operating program?" In order to carry out research activities to answer these questions the following steps must be taken.

1. Identify the expected outcomes of the program, that is, the goals that it was intended to accomplish.

2. Try to identify some of the unanticipated consequences of the program. This is made easier if the evaluator has a good understanding of the interrelations among the organizational variables which are to be changed and those which are not to be changed. For example, it should not be forgotten that successful job enlargement for subordinates automatically implies a change in the role of the supervisor. Experienced change agents and evaluators would be aware of such a simple interrelationship and take it into account in both designing and evaluating the change program.

3. Develop or adopt appropriate measures to ensure that the desired outcome had occurred.

4. Develop or adopt analytic techniques to ascertain whether any effects observed could be attributed to the change program itself rather than to other causes. This involves the use, where possible, of experimental or quasi-experimental designs for evaluation. We shall discuss these fully later in this booklet.

5. Develop commitment of the change agent and the organization (managers, employees and union leaders) to the evaluation measures used in doing the evaluation.

6. Identify and record the costs and likely benefits of a program.

Identification of the Expected Outcomes

This follows logically, from the initial diagnosis of the organization. The diagnosis will identify areas in which change is needed. These needed changes should be explicitly stated as desirable outcomes in the program design stage. These outcomes may be stated in terms of performance (quantity and quality), satisfaction and associated behaviour (turnover, absence, tardiness), quality of union-management relations and associated industrial relations behaviours (numbers of grievances, speed of solving grievances, strike activity). Employees or their representatives should be involved in both diagnostic stages and in choosing and developing the indicators that are used for summative evaluation purposes.

Choice of Evaluative Indicators

A number of different indicators might be needed for summative evaluation.

1. Organization records of productivity, quality, turnover, absences, grievances, etc., can serve as useful indicators. Care must be taken to ensure that the records are appropriate for the purpose, e.g., separate records of voluntary vs. involuntary turnovers, appropriate level of aggregation for the program (individual, work group or organization wide).

2. Specially designed standardized attitude measures might be administered to guage reactions to the program and its consequences. A number of standardized questionnaires (such as the Michigan Organizational Practices Questionnaire) are commercially available. Before launching into this fairly technical area, most organizations would be wise to obtain the advice of a consultant in order to gather useful data by means of attitude scales and measures.

3. Observations of employee or managerial behaviour can be gathered. This can include activity in the workplace, during work meetings, and even in meetings associated with the implementation of the organizational changes. It is also possible to make observations on a sampling basis. In this procedure, observers make randomly timed visits to the workplace over a period of time and make a record of the activity occurring on their arrival. In this way a reasonably accurate picture of the work activity of a large number of people can be assembled in a relatively economical way.

Ideas for what is to be measured can be derived from the initial expectations of managers, union representatives and employees. Managers might be interested in asking the following questions:

What is the impact on productivity?
What is the impact on scrap and waste costs?
What is the impact on the quality of work?
What is the impact on absenteeism and turnover?
What is the impact on worker grievances and work stoppages?
What is the impact on interdepartmental collaboration?
What is the impact on managerial jobs?
What is the impact on health and safety?

Workers might be interested in asking the following questions:

What is the impact on job security?
What is the impact on income security and income improvement?
What is the impact on work-pace, effort, stress?
What is the impact on hygiene?
What is the impact on work satisfaction?
What is the impact on social life on the job?
What is the impact on off-job life?
What is the impact on career opportunities, general life satisfaction, centrality of work in life interests, etc.?
What is the impact on health and safety?

In addition, union representatives might be interested in the following:

What is the impact on union/management relations?
What is the impact on union/member relations?
What is the impact on the role of the union in other companies?

What is the impact on the role of the local and its affairs?
What is the impact on the union as an organization?

Identification of Unanticipated Consequences

A research group at the University of Michigan has argued strongly that assessment research should use an explicit theoretical organization model. Such a model will guide the evaluator in his search for relevant unanticipated consequences. The model that is advocated by most organization theorists these days is an "open system" model of organizational functioning.[7] Such a perspective has at least three implications for evaluation.

1. Relations between the organization and its environment need to be assessed, and any changes between aspects of the organization and its environment noted. For example: Goodman (1979) found that one of the effects of a program to develop autonomous work groups at Rushton Mines was a reduction in safety violations. This resulted in an improved relationship between the firm and the mine inspector at both state and federal levels.

2. Changes to parts of the organization effected at one stage of the process will result in consequent changes in other parts of the organization. For example, creating more cohesive work groups may result in difficulties for inter-group collaboration.

3. Groups whose behaviour is affected at one stage of the process will in turn affect other groups at some later stage. For example, converting a pay system to one involving payment for skill acquisition (in order to encourage employees to develop many skills) may result in all employees in a unit attaining the highest compensation level. Senior employees in a nearby plant, whose pay is based on seniority, may resent the high pay paid to junior employees in the first plant.

These implications of an open system provide us with some guidelines for examining unanticipated consequences. The list given here is general. Significance in particular situations will have to be judged by the parties to the change and its evaluation.

---

[7] "Open system" implies that all dimensions of an operation must be considered, external as well as internal. All these factors interact - customers, suppliers, the social, political and economic environment.

From the first implication we see that all organizations have to draw on resources from their environment, so questions about changes in relationships to the following should be asked in the evaluation:

Raw Materials: How have relations with suppliers changed as a result of the program? Are higher or lower quality inputs now required? Did the nature of supply of raw materials constrain the choices of job and organizational change? Can such constraints be removed by changing suppliers or specifications?

Labour Market: What is the organization's current profile in the labour market? Did the pool of available labour limit the choices that could be made in terms of job and organizational change? Can these limitations be overcome by changing hiring patterns, or by training the current staff? Has there been an impact on quality or quantity of applicants? What was the effect of general economic conditions on the labour market situation?

Financial Resources: What is the organization's current financial position in terms of debt/equity ratio, current cash flow, current policy about reinvesting profits? Are there difficulties in raising additional capital? To what extent did existing financial constraints limit the choices that could be made in terms of job and organizational change? How has the change program affected the financial community's perception of the organization - is "good will" and "sound management" seen as higher after the change? In the Rushton experiment, for example, Goodman asked mining engineers and financial analysts how much they would be prepared to pay for coal mines with different sets of characteristics, that is, a mine with a multi-skilled underground work force versus a mine where each man had just one or two skills. Goodman found the value placed on the former type of work force was higher than that placed on the latter with a less skilled work force.

Legitimacy and Legal Obligations: What are the organization relations with: local community; local, regional, national government? What legal regulations had to be complied with (e.g., safety)? Did these affect the choices available in the program? Did the changes facilitate or hinder meeting these legal obligations? For example, in some jurisdictions the legislation concerning maximum hours of work make it difficult for employees to obtain the full benefit from a flexible hours of work program because the limits imposed by legislation are narrower than the flexible bands that organizations and employees may wish to adopt. Similar questions may be applicable to the local union: What are the local union's relationships with its membership, and with the union headquarters? Do these constrain

the choices that can be made: for example, the local union may wish to collaborate in a QWL experiment while the headquarters may have a policy of maintaining an adversarial, arms-length relationship with employers.

Customers: What is the current relationship of the organization with its customers? Are there problems with accuracy and speed of delivery of goods or services? Do existing contracts constrain organization or job design? For example, some contracts may specify that quality control and inspection must not be performed by manufacturing employees; this would limit the organization's attempts to enrich jobs by adding task identity, autonomy, and immediate performance feedback.

The implications of an open system model alert us to the importance of understanding how the sequence of changes flows through an organization. After implementation of a change, we expect to observe the following sequence of events: there is a change in individual work attitudes and behaviour; this in turn results in an increase in organizational effectiveness and the perceived quality of working life in the organization.

The different variables that need to be measured at each stage of the change program are presented in Table 2. The relationships among variables at different stages of implementation help us to draw conclusions about whether or not the program is the likely cause of any observed changes.

At each of the two stages, we can examine whether or not the expected changes have occurred. If changes in the second stage occur without changes in individual attitudes and behaviour, or even without a positive answer to whether the change was implemented, we would be cautious about concluding that the change program had its intended effect. We would have to search for alternative explanations of the change. For example, in the Alcan experiment, one would expect that those employees who had learned more skills would be more satisfied with the program. Thus, attitudes toward the program could have been correlated with the number of skills learned. If a positive relationship had been found, this would have supported the interpretation that a skill development program had caused the improved job attitudes.

Table 2

Variables Affecting Success in the Last Two Phases of a QWL Program

Stage	Variables of Importance for Evaluation
1. Design and Implementation	
(a) Start-up and Implementation of Program	Goals of management, union and employees Internal/External Politics Labour/Management Relations Role of Joint Committee Degree of Commitment Suspension of Hostilities Politics of Joint Committee - operating style - role in organization
(b) Adoption of Program	Observations About Whether Change has "Taken" Knowledge of Process Participation in Design Skill Development/Amount of Training
(c) Impact on Attitudes	Job Attitudes
(d) Impact on Behaviours and Outcomes	Job Characteristics Performance Turnover, Absence, etc.
2. Establishment of the Program and Spread to Other Parts of the Organization	Attitudes etc. of New Participants After Time Lag Changes Among Non-Target Groups Support for Program by Union and Management

The second aspect to concern us is the impact of changes in one part of the organization on other parts. This has implications both at the design and at the evaluation stage of any change program. At the design stage, it is important to anticipate the necessity for changes in managerial support systems (e.g., compensation procedures, accounting systems, channels for communication within and outside the organization) in order that these are redesigned to support the major components of the change program (such as job redesign). In the evaluation stage, it is important to measure whether these other changes have taken place and whether additional aspects of the organization's design or functioning may be preventing the change program from having a positive effect on attitudes, behaviour, and ultimately on effectiveness and Quality of Working Life.

This awareness of the interrelations between different groups and different subgroups in the organization leads us to ask the following questions when planning the change:

> How interdependent are the target group and other groups in the organization? Do these proposed changes affect their interdependence? Can compatible and parallel changes be made with these other groups as targets (e.g., job enrichment for employees often changes the supervisor's role drastically)?
>
> What changes in pay levels and types of job activities are anticipated? Will these be acceptable to other groups in the organization?

Analytic Techniques

The identification of change in a set of anticipated and unanticipated outcomes is but a first step in deciding whether a particular organization intervention is responsible for causing that change. The evaluation that must be carried out before deciding that an intervention has actually caused the change is akin to the work of the detective in solving a crime. The evaluator explicitly sets up a set of alternative hypotheses about other factors that might either alone or in combination with the intervention have caused the change and then goes about ruling these out. In this way, one might reasonably conclude that the intervention itself is likely to have been responsible for the damage. The task for an evaluator is to arm oneself against criticism that one or another of these alternative causes might really be responsible for bringing about an observed change.

If other factors alone caused the change then we must conclude that the intervention was not very useful. If other factors in combination with the intervention caused the change then the experiment was successful in the set of conditions that existed when the intervention was carried out. In other words, the intervention only had an effect for that particular group of people; or it had an effect when it was preceded by a particular diagnostic questionnaire; or it

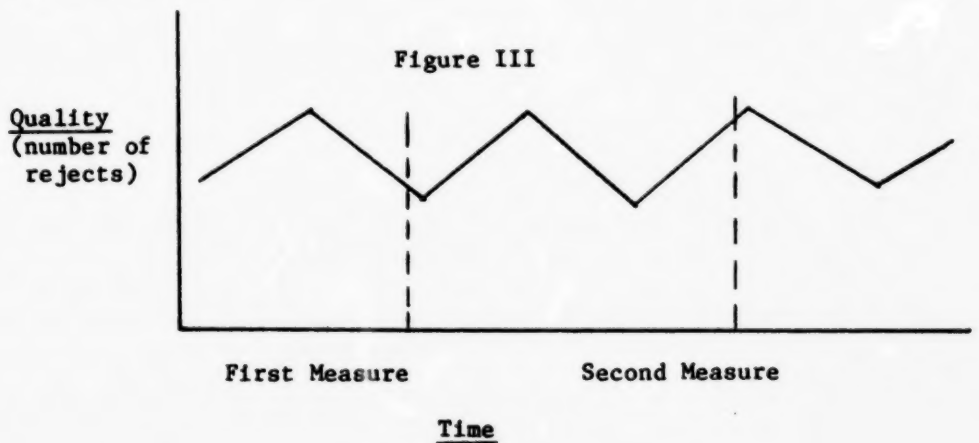
had an effect when it was coupled with other interventions. This specification of the conditions leading to the success of an intervention is the stuff of action research and formative evaluation in which one is trying to maximize the probability of a successful intervention. The difficulty arises when the question is asked: "Okay can we go ahead and implement this change elsewhere in the organization?" The answer is "No". We can only implement the intervention where the ground is fruitful - that is, under the conditions under which the experiment worked. This means that, if the intervention is to be effectively diffused, the necessary conditions have to be created. This may imply preliminary work on other sites to ensure that the conditions are met prior to implementing the program.

It might be useful here to provide an artificial example of the alternative hypotheses that might be entertained about the causal factors involved in a change program. For example, a work group might be assigned to a job that has been enriched in terms of increased responsibility for the members, greater autonomy for workers, and demand for more complex skills. As a result, the quality of work increases and employees respond favorably to a job attitude questionnaire. A critic might argue that these results had little to do with job redesign per se, and instead might argue that six or seven other causes might be responsible, such as:

(1) The volume of work in the unit has declined due to slackening of demand, so workers spend more time on each unit, thus enhancing quality.

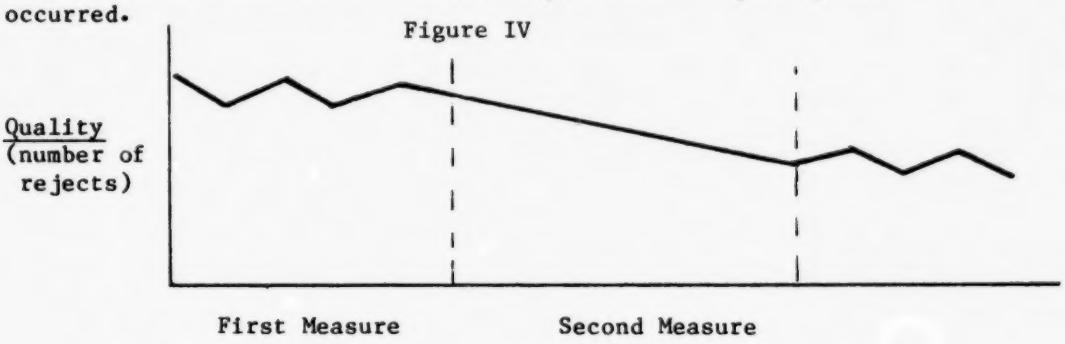
(2) The measurement of work quality typically oscillates around a mean value each month and the first measure just happened to be made at a low point. Thus, the "improvement" observed was no more than the typical monthly fluctuation as shown in the graph on Figure III. As this illustrates, with only the first and second measures one cannot rule out the second explanation. When more extensive time-series of the quality data are examined, the validity of the second hypothesis is supported in this case.

Figure III



In another example (below) the second alternative hypothesis can be ruled out as a stable improvement in quality has occurred.

Figure IV



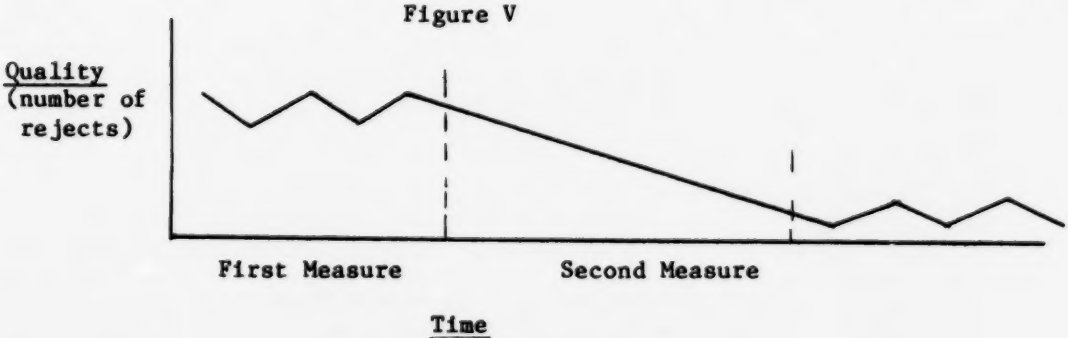
(3) The work group selected for the change might have had some atypical properties:

(a) Initially a work group might have been of very poor quality and therefore could be expected to improve on the second occasion. This could be spontaneous recovery effect.

(b) The group members might have been more technically competent, or more highly motivated than typical work groups in the company and therefore they responded to the change more positively than a typical work group would.

(c) The group members might have been a group of recent hires who were at the bottom of the learning curve for the original job. Thus, the improvement could be attributed to their usual pattern of on-the-job learning, as would be the case in the graph below:

Figure V



Note: It is unlikely that all these work group conditions would operate simultaneously.

(4) The group might have been characterized by a moderate level of employee turnover such that poorer performers leave the group. The observed quality increase could then be attributed to this gradual process of improvement in average quality of worker in the group.

Solutions to the Analytical Problem

There are two ways in which these alternatives hypotheses for explaining the change can be ruled out:

(a) through the design of the evaluation research;

(b) by gathering data relevant to the alternative explanations and showing that these are not potential causes, i.e., that the actual data are not consistent with these other theories. These procedures can demonstrate the inherent implausibility of the alternative explanations leaving us confident that it was the QWL program which was most likely to have caused the observed change. We shall deal with each procedure in turn.

Evaluative Designs

Any evaluation process that we undertake requires making some form of comparison. When we ask, "Did a program have some effect?", we are implying there is a difference within the group before and after the program and that this difference would not exist if the program had not taken place. Either of two evaluation research designs might be adequate to establish that a program had an impact.

(1) Single group design: Here we measure relevant variables prior to the administration of the program and remeasure them at the conclusion. Symbolically the process looks like this.

Observe group at Time 1 ——— Implement Program ———→ Observe group at Time 2

Measure Change: T_2 observation - T_1 observation

This type of evaluation however is susceptible to some of the alternative explanations outlined previously.

(2) Two groups design: One group receives the program, a second does not; both groups are measured after the completion of the program. Symbolically:

Implement Program ⟶ Observe Test Group at Time 2
No Program ⟶ Observe Control Group at Time 2

Change = Observation for Test Group − Observation for Control Group

This design, if implemented properly, provides for complete elimination of the alternative hypotheses mentioned earlier. The problem is that this design is difficult to achieve in organizational practice. Implicitly the design assumes that no differences exist between the groups prior to the implementation of the program. There is one, and only one way of ensuring that there are no initial differences. Random assignment of individuals, work groups, or even organizations to the program and no program conditions is essential. Further, the number of individuals, work groups assigned to the two conditions (programs vs. no program) must be substantial as a rule of thumb 15-20 individuals, etc., need to be assigned to each condition. Clearly, this is a tall order to ask organization experimenters to carry out.

In addition, there are many substantial practical arguments that can be made against random assignment to these two conditions (programs vs. no program).

(1) If one believes that the program is going to have substantial economic or human benefits, then it is hard to justify withholding the program from any potential beneficiaries.

(2) If some people, work groups, or organizations are in trouble, it is unfair to withhold the benefit from them because a flip of the coin assigned them to the no program condition.

(3) There may be difficulty in breaking up an existing organization unit so different sub-units can be allocated to the program/no program conditions.

(4) Only a few units may be available to be assigned to each condition.

These are real constraints on the experimental process. However, in many situations they are not insurmountable. The following strategies for using random assignment can be suggested.

(1) Timing of the application of the program for different individuals, work groups, etc. In most cases, the resources available for program implementation are finite. Thus, not every unit (individual,

group, etc.) can receive the program at the same time. The evaluator should capitalize on this, by randomly assigning units to early and late (in time) versions of the program. Those units receiving the program at a later point in time can provide the comparison (no program) group for those receiving it early (program group).

(2) Even if some units are judged to require the program because they have low morale or low job performance, there will be other units which are on the borderline (they are not very effective, but they are not completely ineffective). In such a case two sets of units should be given the program: (a) those that are in dire need; and (b) a random sample of the groups at the borderline with the remainder acting as comparison groups. The group in dire need should not be formally evaluated in that any improvement could be due to spontaneous recovery. However, a proper evaluation can be carried out by comparing those at the borderline who receive the program with those who do not. This strategy allows for a fair evaluation as well as providing for those who are felt to be in need.

(3) A third argument is that if only a limited number can have the program (due to resource scarcity) then random assignment is the fairest way of assignment.

Of course, these strategies break down when: there are too few units to provide adequate equalization; or when existing work groups cannot be subdivided. In such cases, there are essentially two alternatives:

1. Time series analysis.
2. Non-random comparison groups.

Both of these are second-best as compared with assigning groups at random but are better than nothing at all. Better still would be a combination of the two. The choice depends upon the availability of time and data prior to program introduction.

Time Series Analysis

By using a time series, the group receiving the program acts as its own control. The data obtained prior to the program provide evidence for:

(a) the presence or lack of a learning effect;

(b) the stability or lack of stability of the measures being used;

(c) the influence or lack of influence of other events occurring in the life of the unit;

(d) the presence or absence of a sensitizing effect due to measurement. For example, if an employee is asked to describe a supervisor's leadership style on one occasion, this may affect closeness with which (s)he attends to the leader's behaviour over the course of the succeeding weeks and months. This might influence later responses to a similar questionnaire. A major virtue of time series analysis is that it allows organization members to get used to being measured, thus rendering such measurement unobstrusive and reducing the likelihood of the measurement having an impact on the outcome of the program.

It is through comparing the pattern of data before the intervention with the pattern obtained after that the evaluator can examine the possibility of changes having occurred. Two changes in the data pattern suggest that the intervention may have had an impact:

(a) an increase or decrease in the average level of performance (or satisfaction);

(b) a change in the slope of the time series pattern.

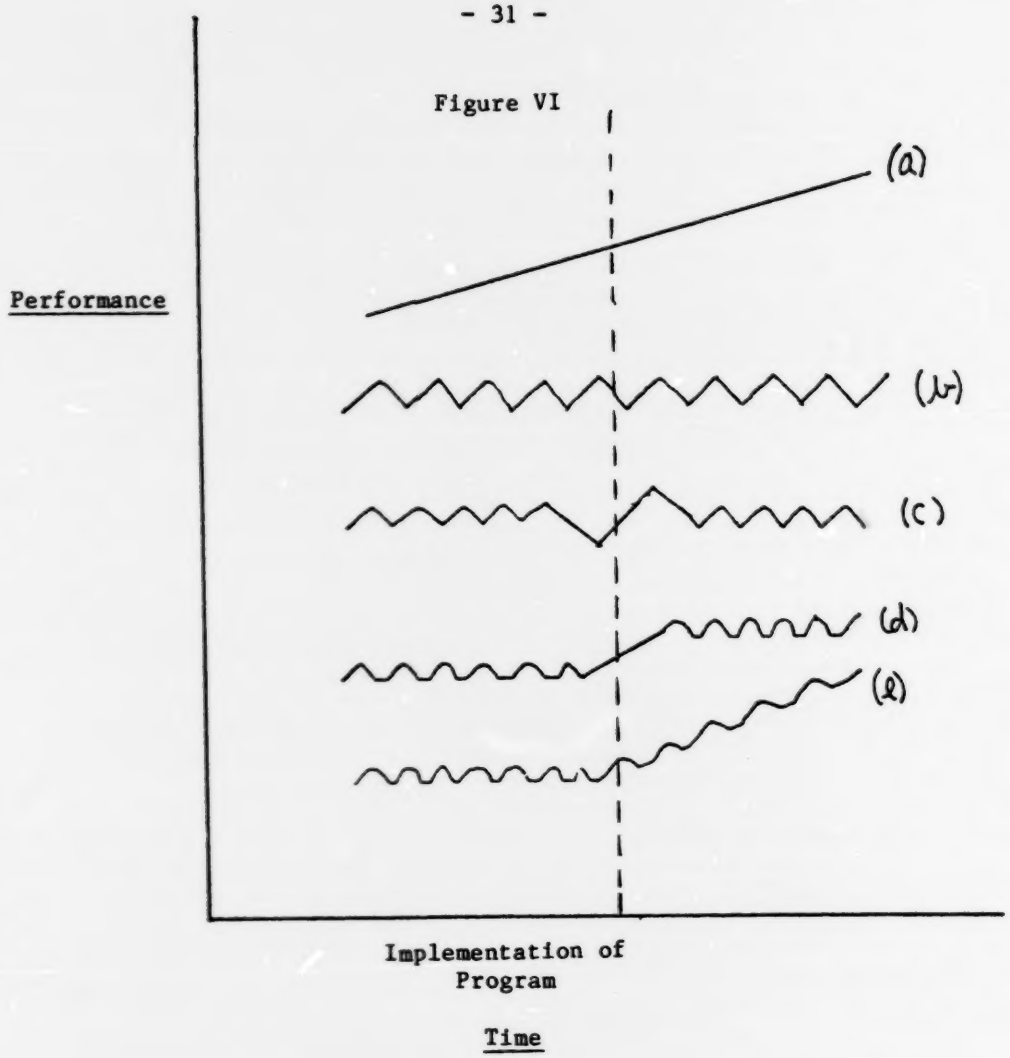
In the diagram below, the intervention was only likely to be responsible for the change in patterns (d) and (e).

Pattern (a) merely suggests a sustained trend (in performance or satisfaction); pattern (b) suggests fluctuation around an unchanged average level; pattern (c) suggests a temporary change (perhaps due to the program) that is not sustained over a longer period of time.

The major difficulty with the time-series design is the possibility that some other event occurred at approximately the same time as the program and was responsible for the change. This alternative can be ruled out by:

(1) Keeping a careful record of the events occurring in the organization and its immediate political and economic environment. If there are no plausible alternative causes, one may justifiably conclude that the intervention caused the observed changes.

Figure VI



(2) Developing parallel measures of variables that you expect to be unaffected by the change. Then, in the experimental group measure those variables that you expect to change and those that you expect not to change at each point in time of the series. After the intervention, the pattern of results should become dissimilar. As an example of this, one might expect that a QWL program will affect voluntary turnover but not involuntary turnover (due to sickness, pregnancy, spouse's job relocation, etc.). Thus, the pattern of these two types of turnover can be observed over a long time period. If the QWL program has had an effect, the voluntary turnover rate should decline, while that for involuntary turnover should remain stable.

In our Alcan example, the major dependent variables were productivity and job attitudes. Each of these was measured simply before and after the implementation of the experiment.

1965 - 1968	No change
1968 - 1969 (mid)	Increased productivity due to technical improvements
1970 - 1971	12% increase in productivity coupled with no technical change. However, foremen judged that only about half of this gain (7%) was due to the new organizational arrangements.

We can see that the data do exist for a time series analysis. However, the analyses are confounded by the fact that we are dealing with two sequential changes: (a) the technical change of 1968-69, and (b) the organizational changes of 1970-71. In this case, it would probably be useful to examine the data based on quarterly production figures (unless these were seasonal production cycles). This would provide us with sufficient data to look for a flattening of the learning curve following the technological changes and the increase in productivity assumed to be due to the new organizational arrangements. If this flattening were not observed, we might conclude that the organizational changes had little impact and that the productivity increases were merely a continuation of those caused by the technological improvements (see pattern (b)). The first graph (Figure VII) shows two discontinuities induced, one by technical changes, the other by organizational changes. The second graph (Figure VIII) shows a single discontinuity that is caused by technological change.

Figure VII

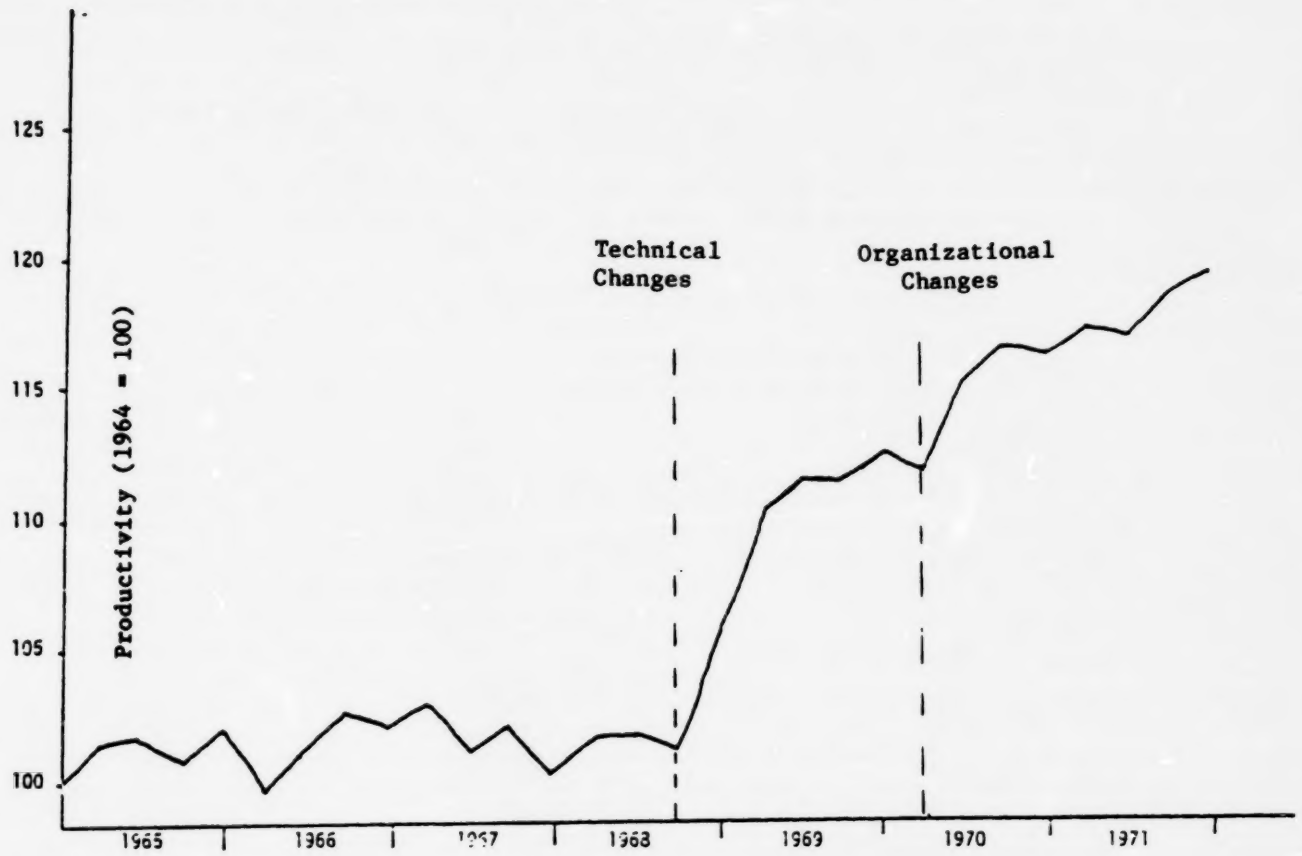


Figure b) Pattern of productivity data to be expected if technical change alone had an effect.

Figure VIII

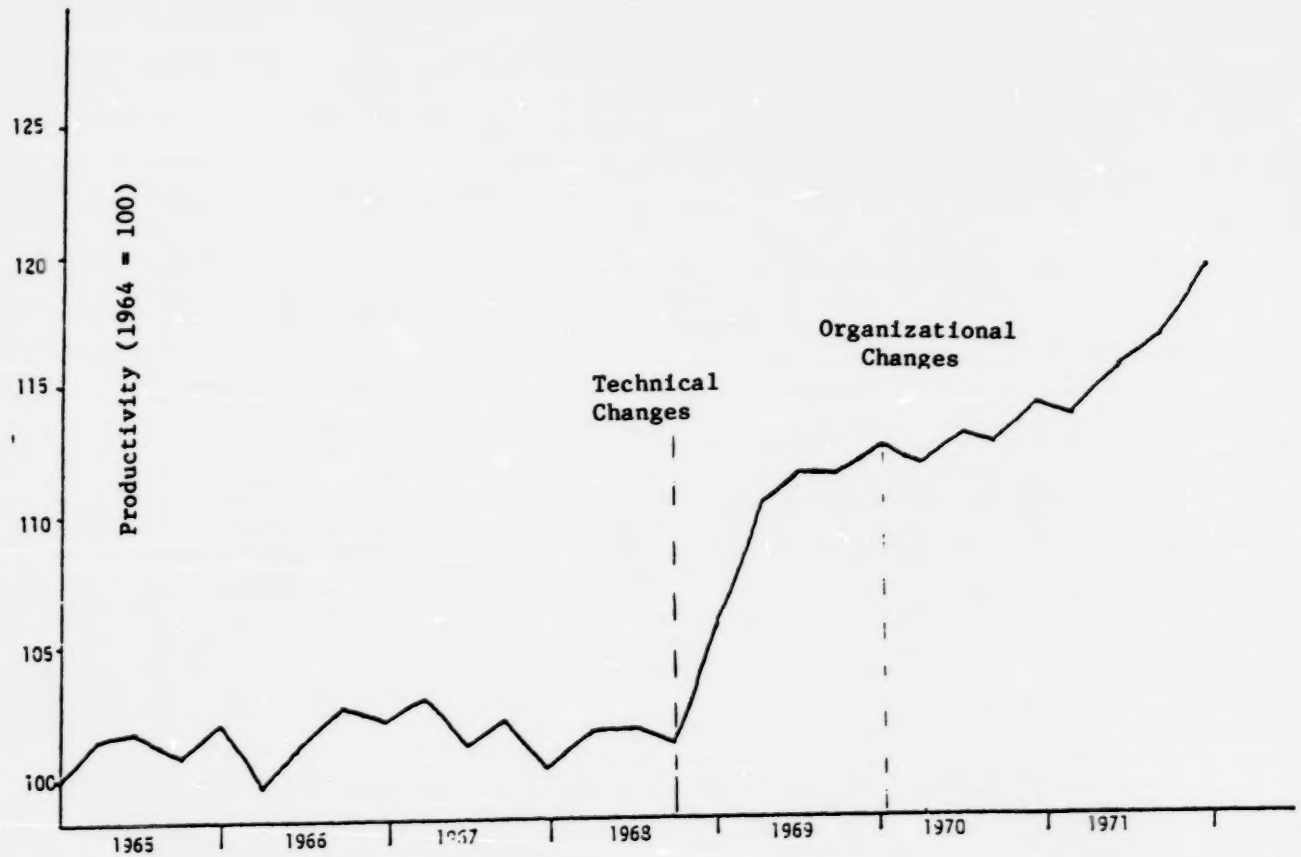


Figure a) Pattern of productivity data to be expected if technical and organizational chang had an effect.

The discussion of the examination of job attitudes will be postponed until we have discussed the second analytic framework.

Non-random Comparison Group

In this type of analysis two groups (of individuals, work groups, or organizations) which are not formed randomly are compared before and after the program has been implemented in one of the groups. This type of design provides quite good control for most of the problems we have discussed so far.

1. Both program and no program groups experience the same sequence of extraneous historical events.

2. When the groups are found to be as similar as possible prior to the program, the threat that initial differences caused the change is minimized. This means that they should be similar not only in terms of interest in the experiment (productivity, job satisfaction, turnover, absence, etc.), but also in terms of the way they were recruited to take part in the study. Comparing a group of volunteers who take part in a QWL program to a group who were disinterested in such a program leaves the evaluator open to the charge that observed changes might be due to motivational differences between the groups rather than due to the program effects themselves. Similarity in terms of interest will help to rule out the possibility that the group undergoing the change program has improved as a result of a spontaneous recovery from an initially disadvantageous position.

It is possible, of course, that being tested prior to the program sensitizes people to the content of the program so that the observed change is due to pre-testing and the program change in combination. One way to deal with this problem is that pre-testing always occurs when implementing the program elsewhere. This is probably a good idea anyway, the diffusion of any QWL program should be preceeded by its own diagnostic phase which can include this pre-measurement. The second way to deal with this is: (a) use only measures which are non-reactive such as records of turnover, production, etc.; (b) develop a time-series design so that employees become used to attitude questionnaires or observation so that they cease to react to or be sensitized by the measuring techniques used; and (c) extend to four groups - two with treatment but one with and one without pre-measurement; two without treatment but one with and one without pre-measurement.

For example, at Alcan we find the following evaluation of job attitudes.

Job Attitudes	Pre-Measure:	September 1970 (Questionnaire)
	Post-Measure:	October 1971 (Questionnaire)
	Post-Measure:	October 1971 (Interview)

The differences in the questionnaire data were not significant. This reflects the difficulty in obtaining a sensitive instrument. A useful supplementary analysis would have been to examine the changes for those who responded positively to the experiment (75 per cent) compared with those who were indifferent or negatively disposed (25 per cent).

The interview data suggested that there were increases in:

felt recognition for skill and performance;
collaboration within the group;
increased group motivation;
increased importance of work;
increased interest in work.

In addition, most employees felt positively toward the job enlargement program. These interview data lack the element of comparability that we have deemed essential to any evaluation exercise. As a result, it is hard to know whether these data reflect true changes brought about by the experiment or whether they would have occurred anyway. Given the fact that the plant contained several other similar work units, it should have been possible to identify two or three work units which had not undergone the change and carried out post-experiment interviews with their members to provide a rough comparison of the job attitudes after the experiment. It would have been still more useful to interview half the members of the experimental group and half the members of a comparison group than to interview all members of the experimental group.

One problem remains. This concerns the comparability of any comparison group. We are told that the experimental group had "a widely fluctuating work load" and that "crew organization continually shifted and foremen would have to supervise 25 to 45 workers on each shift, well above the average of 16 workers per foreman." In addition, the group occupied twice as much floor space as other departments. It seems that an additional change (above and beyond the job redesign) was the reduction of the foreman's span of control by about "8 to 16 workers."

This muddies our analysis in two ways:

(1) The change involved a reduction in group size as well as job redesign.

(2) The experimental group started off in a disadvantaged state in comparison to the other groups.

It may therefore only be possible to generalize the observed improvements to situations that are similar. As there are probably many such situations in organizations, the lessons from this change attempt to

have practical implications for the manager. We have presented here a rather elaborate procedure of evaluation. We have done this deliberately. We do not believe that there are easy answers to the kinds of questions that a manager might raise:

> How can I get a quick and simple assessment that my change program is working?
>
> How can I convince management to extend the experiment to other organizational units?

Any data that we might generate to answer these questions are likely to be confounded by the factors outlined earlier. The only way to answer such questions is through the careful evaluation procedures outlined above. We would emphasize that, with some thought and creativity, a good evaluation design is no more costly or more difficult to implement than an inadequate one.

Finally, perfection is difficult to achieve. Evaluators should simply do the best they can in: (a) assessing the likely threats to the validity of their evaluation procedures; and (b) designing evaluations to reduce the probability of this threat occurring.

VI. Action Research and Formative Evaluation - A Feasible Approach

Action research, a well established technique for implementing a planned change lends itself well to formative evaluation. Two key features are the presence of someone in a consulting role and the use of feedback data as the prime force in effecting change. A third feature of action research is that the activity is oriented toward problem-solving as a means of implementing change. In other words, the consultant shares the concern of the client group (project team) in trying to solve the problem of achieving the change. The difference between the consultant role and the project team role[8] is that the consultant acts as an independent observer and helper to the team or the group trying to implement the change. The project team concerns itself with taking direct action to implement the change and working directly with the people at the work-site.

Action research can also be thought of as an incremental process of change which can go through the following steps. It will be useful to trace the various stages highlighting the role of evaluation.

---

[8]The term "role" is used because the consultant can be someone external or internal to the organization.

1. Problem Diagnosis - A problem may be perceived and diagnosed by people in the organization by themselves or with the aid of a consultant. Data gathering at this stage is concerned with finding out the magnitude of problems, sources of problems, and consequences of problems.

2. Goal Setting - In this stage, the consultant works with the project team in the organization to define the objectives, target, ambitions, and aspirations for their change program. The project team should concern itself with three factors at this stage.

Definition of objectives - The more precisely the objectives are defined before a project is begun, the more that thinking is clarified and the evaluation task made easier.

Definition of evaluation criteria - Considerable attention should be paid to trying to define how achievement of an objective will be recognized. In other words, if an objective can be defined both as an aspiration and as a measurable level of performance, both the implementation task and the evaluation task will be made easier (i.e., you will know more clearly what you are trying to achieve and you will be able to identify more clearly when you have achieved it).

Planning of the change program - The work devoted to goal definition will greatly aid the planning of the various steps in the change program itself. Identification of evaluation criteria will also give more direction to various aspects of the change program. For example, it is much easier to plan and structure a program on supervisory leadership styles if you say you want "to achieve a 9, 9 style of leadership among 80 per cent of managers" than it is to have a program "to improve managers' leaderships styles."

In all of the foregoing stages the consultant should work with the project team as both an informational resource and as an aid to the communication and decision-making processes of the group. This consultant role is called process consultation.

3. First Trial of Program - When the project team begins to introduce the changes that are necessary to effect program implementation, the consultant can have either or both of two roles. One role is to be actually involved in the change process, perhaps acting as a trainer, counsellor, interviewer, mediator, etc. The second role, which is crucial to action research, is to gather data on the interim results of the program.

Two kinds of data are involved: data on how people are responding to the training, educating and administrative aspects of the program (reaction evaluation); and data on how people are responding to the content of the training, etc. (evaluations of learning, attitude and behaviour changes).

Reaction evaluation can be concerned with such aspects as how much people liked a trainer, the program environment, the lodgings, and food. Negative reactions to any of these aspects may jeopardize acceptance of the program content, whereas positive reactions may only assure that people will be tolerant of such content.

Content evaluations are the more important evaluations because these try to assess whether people are likely to respond in a significant manner to the program, by changing their attitudes and behaviours. For example, a person may say, "Well I sure had a good time at that seminar" (positive reaction) "but it would never work in our organization" (negative content) or conversely, "It was kind of a slow program" (negative reaction) "but there is a lot of relevance in it for our organization" (positive content). A positive response on content is likely to have a lot more impact on behaviour than a positive response on reactions to programs. Obviously, positive reactions on both types of evaluation offers the most promise.

The type of evaluation which is concerned with interim measures of progress, both of techniques and of content, has been previously described as FORMATIVE evaluation. Formative evaluation in action research is used to measure progress toward a goal as well as assessing the effectiveness of various techniques. For example, an implementation program may involve a one-week retreat for training purposes with the result that most people spend all their time on social activities. This would likely generate a positive reaction evaluation but a much less favourable response on content evaluation. Based on this experience, the project team might try a different training strategy the next time around for another group of trainees. Both technique as well as results have thus been assessed in this type of formative evaluation.

The initial data gathering in the diagnosis stage may give the consultant or the project team some opportunities to do "before and after" formative evaluations. Otherwise, most formative evaluations are concerned with degree of goal achievement or else deal with qualitative results. This is because it is often difficult to obtain quantitative data on a program until long after the program is under way. Yet, it is often necessary to have interim data to see whether the program is on the right track or whether the techniques used are appropriate for the program participants.

4. Initial Feedback of Results - The consultant in this stage feeds back the interim or formative evaluations to the project team and then participates in the discussions on whether the program is still on target or not. Depending upon the nature of the feedback, the program may proceed as planned, may have to be modified in some way or more problem-solving may have to be done. In a sense, the initial feedback stage is similar to the initial diagnosis stage in that the consultant and the project team are looking for data relevant to a problem and trying to decide on a course of action. During the

change-over project at Alcan, for example, bi-weekly meetings lasting an hour and a half were held between the foremen and the internal consultants. In addition to being training sessions, these meetings were also used to discuss any problems that had arisen, evaluate alternatives and recommend corrective action. Focal issues discussed in earlier meetings were brought back for further evaluation in the light of new experience gained. Several changes had not turned out to be functioning as planned, others announced at the outset were never realized. Continual monitoring evaluations, provided the basis on which corrective actions could be implemented. The beneficial results achieved through this joint problem-solving approach persuaded the parties, as noted earlier, to implement a similar system for all employees in the casting department.

5. Action Planning - Based on the results of the feedback, the steering committee decides on a series of further goals and actions (which may be a reaffirmation of previous goals and actions), assigns responsibilities, timetables and budgets to the project team and agrees on a further feedback meeting at some future date.

6. Steps 3, 4 and 5 are repeated as necessary until the feedback eventually begins to show that the objectives have been achieved or that the original problem(s) are no longer a source of concern. The number of times that these steps are repeated will depend on the magnitude of the problems and available time and resources. When the new program has been effectively implemented, the consultant in action research is ready to undertake the final evaluations.

7. Summative Evaluation - This type of evaluation has been described as being concerned with trying to measure or assess the overall results of a program, whether it achieved the original objectives, how much the program cost, how much the program saved, unanticipated consequences, and so on. Summative evaluation was more fully described in a previous section of this booklet and so it is sufficient to say here that summative evaluation is appropriate for the completion of the action research process.

In summary, action research is considered to be a data-based method for introducing, monitoring, aiding and evaluating organization change. It is especially useful for the introduction of QWL programs and provides a technique to both implement and evaluate a project simultaneously. While action research is most often done with the aid of an independent consultant, the action research role can be performed by someone within the organization as well. What the technique requires, more than sophistication of data gathering and process consultation, is openness among people in the organization to ongoing evaluation and willingness to learn from the data. To quote some well-known writers in the field, action research can be embodied in two modes of thinking, "Action should follow research...and research should follow action."[9]

---

[9]French and Bell, Organization Development, 1978.

VII. Conclusion

The introduction to this booklet stressed the importance of planning for evaluation and trying to treat evaluation procedures as an integral part of the implementation of a QWL program. The balance of the booklet described several approaches, techniques and cautions for evaluation. While the subject may have seemed rather complicated at times, our major conclusion is to stress the importance of planning for evaluation. Thinking about how to evaluate the results of a program should be done almost simultaneously with thinking about how to implement the program. Poor evaluation can produce a conclusion that a program had failed when it had actually succeeded or that a program had succeeded when it actually failed. With the first conclusion, an organization might fail to support a valuable program; with the other conclusion, an organization might waste a lot of resources promoting a useless program.

When planning for evaluation, it is also useful to sort out who will do what, or who has the various roles of evaluator, implementor and recipient of program changes. It may not always be possible to have separate people perform these various roles but it is highly advisable to recognize the different obligations and responsibilities associated with each. Consciously recognizing these differences allows greater concentration on the tasks of both implementation and evaluation which hopefully will encourage better quality performance in both roles.

Planning ahead for evaluation allows more opportunities to deal with issues of contamination of evaluation designs or to try to use more of a quasi-experimental approach to evaluation. Even if one has the implementation view of trying to force the desired results of a program, these actions will be more effective if they are based on past results or experiences that are reasonably valid. The more one tries to control for extraneous causes of change in assessing the impact of a program, the more one can use techniques and strategies from earlier programs with confidence in future programs.

There is a continuing trade-off between formative and summative evaluation. The implementation approach, geared to getting results is most closely allied with formative evaluation; the scientific view of adding to knowledge is most closely associated with summative evaluation. Yet, it is misleading to suggest that these two positions are extremes without any common ground. The project team will still have an interest in the summative picture once a program is finally implemented (unless they are willing to accept the overall value of a program on faith alone). Evaluators have an interest in the implementation question as it is of little use to know what overall results were achieved if one could not duplicate the experiment or learn how the results were achieved. Action research was suggested as a feasible approach to provide a common meeting ground for both the implementation and evaluation perspectives. As with any compromise,

action research does not maximize the preferences of either of the first two approaches but it does have the great advantage of being a well-established, successful implementation technique for QWL programs.

Finally, two other points should be highlighted in our conclusion. The first is that far more attention should be given to assessing the costs and payoffs from QWL programs in terms of real and estimated dollars. Organization managers pay serious attention to costs of practically any aspect of hardware in an organization and frequently examine the payoff from investments in hardware, yet rarely does one see similar examinations of costs and payoffs data from investments in human resources. Admittedly, the payoff data can be difficult to obtain in some cases, but it is surprising how little recording or accounting is done regarding the costs of QWL programs. One might argue as Rensis Likert[10] did years ago, that human resource programs will not achieve the same influence in organizations as other more traditional functional programs unless human resources investments can be talked about in the same dollars and cents language as capital investments.

The second point is that organizations which implement and evaluate QWL programs should be willing to share their knowledge gained with other organizations. Most published knowledge about the results of QWL programs comes from a few academic studies and these represent a small percentage of such activities in the area. Quite possibly, a lot of potentially valuable knowledge is lost simply because few people are aware of the successes and failures of other organizations' programs. Of course, few people want to talk about their failures or lack of successes but the knowledge gained from these can often be of immense value to others. For example, the follow-up results of the QWL experiments at the General Goods Topeka plant or the Non-Linear Systems Company have contributed a great deal to a more realistic concept of the applicability of QWL and organizational innovation programs. A number of European countries have tripartite research bodies from government, industry (including labour unions) and universities to sponsor the involvement of action researchers in QWL projects. In the short term, the organizations get the benefits of more effective program implementation as a result of the involvement of the action researchers. In the long term, both industry and academic knowledge benefit from the results of research on the program published by the action researchers. Such programs seem to be an effective way of making progress on the optimal utilization of human resources. While not everything from European initiatives in human resource management can be applied in Canada, this commitment to evaluation and research by industry (management and unions), government and the universities does seem to be a highly appropriate direction to follow.

---

[10]Likert, R.

References

Agnew, N.W. and Pyke, S.W., The Science Game: An Introduction to Research in the Behavioural Sciences, Toronto, Prentice-Hall, 1969.

Beckhard, R. and Harris, R.T., Organizational Transitions: Managing Complex Change, Don Mills, Ontario, Addison-Wesley, 1977.

French, W.L. and Bell, C.H., Organization Development 2nd Edition, Toronto, Prentice-Hall, 1978.

Jones, J.T., Achieving Joint Organizational, Technical and Personal Needs: The Case of the Sheltered Experiment of an Aluminum Casting Team, in Davis, L. (Ed), The Quality of Working Life: Volume Two-Cases and Commentary, New York, The Free Press, 1975.

Johnston, C.P., Alexander, M. and Robin, J. Quality of Working Life: The Idea and Its Application, Ottawa, Labour Canada, September, 1978.

Likert, R., The Human Organization, Toronto, McGraw-Hill, 1967.

Nadler, D.A., Feedback and Organization Development: Using Data Based Methods, Don Mills, Ontario, Addison-Wesley, 1977.

Ondrack, D.A. and Evans, M.G., The Quality of Work Life Program at Petrosar, Ottawa, Labour Canada, 1980.

Suggestions for Further Reading

Armenakis, A.A. and Feild, H.S. Evaluation of organizational change using non-independent criterion measures. Personnel Psychology, 1975, 28,39-44.

This short note demonstrates the use of time-series analysis and the appropriate statistical analysis. Very clearly written.

Armenakis, A.A. and Zmud, R.W. Interpreting the measurement of change in organizational research. Personnel Psychology, 1979, 32,709-723.

The authors develop a procedure for controlling observed change due to scale recalibration. In addition the ideas of Campbell and Stanley are applied to help understand whether or not organizational change occurred.

Campbell, D.T. and Stanley, J.C. Experimental and Quasi-experimental Designs for Research. Chicago, Ill.: Rand-McNally, 1963.

The book for anyone seriously engaged in evaluative research. Sets out as clearly as possible the alternative hypotheses that might account for change and develops as rigorously as possible designs for ruling out these hypotheses. On the whole it is written rigorously but clearly.

Cook, T.D. and Campbell, D.T. The design and conduct of quasi-experiments and true-experiments in field settings. In Dunnette, M.D. (Ed.) Handbook of Industrial and Organizational Psychology. Chicago, Ill.: Rand-McNally, 1976.

An up-date of the Campbell and Stanley book. Again clearly written, but quite technical. One strength of the article is an attempt to identify organizational situations in which it should be possible to perform true experiments.

Evans, M.G. Opportunistic organizational research: the role of patch-up designs. Academy of Management Journal, 1976, 18,98-108.

Applies the ideas of Campbell and Stanley about the alternative hypotheses that could account for observed changes. Shows how a good quasi-experimental design may be built up out of a parallel series of designs each of which contain major flaws.

Goodman, P. Assessing Organizational Change: the Rushton Quality of Work Experiment. New York, Wiley, 1979.

Provides a running description and evaluation of a recent quality of working life experiment. Uses economic and attitudinal data in the evaluation with a very sophisticated evaluation design. Very clearly written.

Koch, J.L. and Rhodes, S.R. Problems with reactive instruments in field research. Journal of Applied Behavioural Science, 1979, 15, 485-506.

This illustrates ways of identifying and controlling for reactive measurement and scale recalibration in evaluative research.

Mintzberg, H. The Nature of Managerial Work. New York: Harper-Row, 1973.

The appendices to this book illustrate in detail the strengths and weaknesses of participant observation as a data-gathering technique.

Morris, L.L. Program Evaluation Kit. Beverly Hills, Calif.: Sage Publication, 1978.

This consists of a series of eight booklets:

Evaluator's Handbook
How to Deal with Goals and Objectives
How to Design a Program Evaluation
How to Measure Program Implementation
How to Measure Achievement
How to Measure Attitudes
How to Evaluate Statistics
How to Present an Evaluation Report

While these books are aimed at the evaluation of educational programs within school settings, they are full of sound advice for the novice evaluator. Each is written in a relatively simple step-by-step fashion and are designed to provide enough information so that the evaluator can do the best possible in organizational situations. They are well worth reading as elaborations of the procedures recommended in this Handbook.

Robson, C. Experiment, Design and Statistics in Psychology. Harmondsworth, U.K.: Penguin, 1973.

This is a straightforward and simply explained introduction to some of the statistics needed to conduct evaluation research.

Seashore, S.E., Lawler, E.E., Mirvis, P.H. and Cannon, C. (Eds.). Observing and Measuring Organizational Change: A Guide to Field Practice. New York: Wiley, in press.

When this appears it will be the practitioners handbook. It represents the culmination of six years experience in evaluating organizational change. It points up the issues to be aware of at each stage of the evaluation research. It emphasizes the importance of flexibility in the evaluation process. Our handbook has benefited from our reading of a pre-publication copy of this book. If you are going to be involved in evaluation research, read this book!

Taylor, J.C. and Bowers, D. Survey of Organizations. Ann Arbor, Michigan: Institute for Social Research, 1972.

A handbook of questionnaires designed to tap a variety of organizational characteristics. The culmination of 20 years of organizational diagnosis by the Michigan researchers.

Tchobanian, R. Unions and the humanization of work. International Labour Review, 1975, 3,

Lists the issues labour is concerned with in a QWL program.

Webb, E.J., Campbell, D.T., Schwartz, R.D. and Sechrist, L. Unobstrusive Measures: Non-reactive Research in the Social Sciences. Chicago, Ill.: Rand-McNally, 1966.

This describes a variety of opportunities that are available for taking measures of human behaviour without the subjects being aware of the measures. Most useful for organizational research are the sections on archival information and simple observations. The strength of the book is in the emphasis on the use of the types of data to supplement and compare with the question and interview data usually gathered. The book also raises the ethical issues inherent in such research.